

The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans

Mike Cross, Rudo Kieft, Robert Sabatini, Matthias Wilm¹, Martin de Kort², Gijs A. van der Marel², Jacques H. van Boom², Fred van Leeuwen³ and Piet Borst⁴

Division of Molecular Biology and Centre of Biomedical Genetics, The Netherlands Cancer Institute, 1066 CX Amsterdam, ²Leiden Institute of Chemistry, Gorlaeus Laboratories, 2300 RA Leiden, The Netherlands and ¹EMBL, Protein and Peptide Group, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

³Present address: Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA 98109-1024, USA

⁴Corresponding author
e-mail: pborst@nki.nl

DNA from Kinetoplastida contains the unusual modified base β -D-glucosyl(hydroxymethyl)uracil, called J. Base J is found predominantly in repetitive DNA and correlates with epigenetic silencing of telomeric variant surface glycoprotein genes in *Trypanosoma brucei*. We have now identified a protein in nuclear extracts of bloodstream stage *T. brucei* that binds specifically to J-containing duplex DNA. J-specific DNA binding was also observed with extracts from the kinetoplastids *Crithidia fasciculata* and *Leishmania tarentolae*. We purified the 90 kDa *C. fasciculata* J-binding protein 50 000-fold and cloned the corresponding gene from *C. fasciculata*, *T. brucei* and *L. tarentolae*. Recombinant proteins expressed in *Escherichia coli* demonstrated J-specific DNA binding. The J-binding proteins show 43–63% identity and are unlike any known protein. The discovery of a J-binding protein suggests that J, like methylated cytosine in higher eukaryotes, functions via a protein intermediate.

Keywords: antigenic variation/DNA-binding protein/DNA modification/epigenetic silencing/VSG

Introduction

DNA modification is a feature of many prokaryotic and eukaryotic genomes. Cytosine methylation, a covalent modification of DNA at CpG dinucleotides, is the most prevalent and best studied DNA modification in eukaryotes. It is involved in gene silencing at the level of transcription initiation and elongation, it can act as an imprinting signal and it has been suggested to be required for the suppression of intragenomic parasitic sequences such as retroviruses, *Alu* elements and transposons (reviewed in Jaenisch, 1997; Kass *et al.*, 1997; Yoder *et al.*, 1997; Colot and Rossignol, 1999).

Although kinetoplastid flagellates such as the parasite *Trypanosoma brucei* contain no methylated cytosine, they

do have β -D-glucosyl(hydroxymethyl)uracil (called J), a modified base that replaces a fraction of thymine in their DNA (Gommers-Ampt *et al.*, 1993; Borst and Van Leeuwen, 1997; Van Leeuwen *et al.*, 1998c). J is made in two steps, thymine first being converted to 5-(hydroxymethyl)uracil, which is subsequently glucosylated (Van Leeuwen *et al.*, 1998b). Whilst J is enriched in the telomeric repeats of all Kinetoplastida, in *T. brucei* it is also found in other repetitive sequences and in the telomeric variant surface glycoprotein (VSG) gene expression sites involved in antigenic variation (Van Leeuwen *et al.*, 1996, 1997). Here J is present in genes in silent sites but not in the active site (Bernards *et al.*, 1984; Pays *et al.*, 1984; Van Leeuwen *et al.*, 1997). Recent experiments have shown that a decrease in the level of J in the genome can increase transcription near a silent VSG promoter (Van Leeuwen *et al.*, 1998b). These data support the hypothesis that J is involved in the transcriptional repression of silent VSG gene expression sites, perhaps in a manner akin to methylated cytosine in higher eukaryotes.

J is conserved across the Kinetoplastida (Van Leeuwen *et al.*, 1998c), most of which do not undergo antigenic variation. In *T. brucei*, J is abundant in and around non-transcribed repetitive sequences, and clustering of J in other Kinetoplastida has also been observed. It is therefore possible that the primary function of J is not in regulating transcription of specific genes, but rather in suppressing recombination between repetitive DNA elements to maintain chromosome stability. Indeed, we have found recently that DNA rearrangements sometimes observed with VSG gene expression site switching in *T. brucei* are reduced when the level of J in the genome is increased (Van Leeuwen *et al.*, 1998b). To accommodate the different putative roles of J in transcriptional repression and maintenance of genome integrity, we have proposed that J results in the formation or stabilization of a condensed chromatin structure (Van Leeuwen *et al.*, 1998b,c). This is likely to occur through the recruitment of specific proteins that modify chromatin, as is the case for methylated cytosine in mammals, where proteins such as MeCP2 bind and mediate transcriptional repression (Meehan *et al.*, 1989; Lewis *et al.*, 1992; Hendrich and Bird, 1998). Another example is silencing in the yeast *Saccharomyces cerevisiae* (reviewed in Grunstein, 1997; Sherman and Pillus, 1997). Here a complex of silencing proteins is recruited to telomeric repeats and to the mating type loci.

Here we describe the identification of J-binding proteins from *T. brucei* and the related kinetoplastid parasites *Crithidia fasciculata* and *Leishmania tarentolae* that specifically bind double-stranded DNA containing J. Following purification of the *C. fasciculata* J-binding protein, we identified the corresponding gene and then cloned genes for similar proteins from *T. brucei* and *L. tarentolae*. The J-binding protein is conserved between

Table I. Effect on J-specific complex formation of oligonucleotide duplexes and other competitors used in this study^a

Competitor	Sequence ^b	Inhibition of binding to J-containing DNA
tel-J	5'-TAGGGTTAGGGTTAGGGTTAGGGT-3' 3'-AJCCCAAJCCCAAJCCCAAJCCCA-5'	+
tel	5'-TAGGGTTAGGGTTAGGGTTAGGGT-3' 3'-ATCCCAATCCCAATCCCAATCCCA-5'	-
tel-V	5'-TAGGGTTAGGGTTAGGGTTAGGGT-3' 3'-AVCCCAAVCCCAAVCCCAAVCCCA-5'	-
vsg-J	5'-CAGAAGGCAGCJGCAACAAG-3' 3'-GTCTTCGJCGACGTTGTTC-5'	+
vsg	5'-CAGAAGGCAGCTGCAACAAG-3' 3'-GTCTTCGTCGACGTTGTTC-5'	-
bio-vsg-J	5'-CAGAAGGCAGCJGCAACAAG-3' bio-3'-GTCTTCGTCGACGTTGTTC-5'	NT
<i>E.coli</i> DNA		-
Calf thymus DNA		-
Phage ϕ e DNA	contains 5-(hydroxymethyl)uracil	-
pdJ		-
UDPG		-
GDPG		-

^aDNAs were used as unlabelled competitors for binding to probes tel-J or vsg-J. +, competition; -, no competition; NT, not tested.

^bSequence of duplex oligonucleotides or relevant features of the DNA. J, β -D-glucosyl(hydroxymethyl)uracil; V, 5-(hydroxymethyl)uracil; bio, biotin.

these three species, but shows no significant homology with any other known protein.

Results

A JBP activity in trypanosome nuclear extract

We detected J-specific DNA-binding proteins in nuclear extracts of *T.brucei* bloodstream form using a bandshift assay with chemically synthesized DNA oligonucleotides of different sequences that contain base J (Van Leeuwen *et al.*, 1996, 1997, 1998a; see Table I): 4 \times telomere repeats (tel-J) and a VSG gene sequence containing a *Pvu*II restriction site (vsg-J). To control for non-specific binding, probes of the same sequence but with Thy in place of J were used. J-specific binding was observed only with double-stranded versions of the probes. Figure 1A shows that upon addition of extract to the reaction with probe tel-J, several complexes were formed. The slowest migrating complex is likely to be J-specific as it was absent from the reaction with probe tel, the J-free version of tel-J. This was verified by competition experiments (Figure 1A; summarized in Table I). A 27-fold molar excess of unlabelled double-stranded tel-J abolished formation of the upper complex, as did double-stranded vsg-J; vsg-J did not compete complex formation as effectively as tel-J, but this was difficult to quantify due to the low abundance of the J-specific complex. The J-free versions of these DNAs showed no significant competition, neither did *Escherichia coli* or calf thymus double-stranded DNA. The J-specific complex was observed when vsg-J was used as probe, and the effect of different competitor DNAs on complex formation with vsg-J was the same as that seen with tel-J (data not shown). None of the additional non-J-specific complexes observed in the bandshift assay were investigated any further.

Crithidia also have base J, so we repeated the binding studies with nuclear extract prepared from *C.fasciculata*

(Figure 1B). A J-specific bandshift with probe tel-J was again observed, although compared with *T.brucei* the *C.fasciculata* complex migrated much further in the gel, and it was more abundant (see also Figure 2A) and less sensitive to the assay conditions. Complex formation was competed effectively only by J-containing DNAs. In this case, vsg-J competed ~5-fold less than tel-J. Binding reactions with *C.fasciculata* extract and probe vsg-J also produced the J-specific complex (data not shown). The slower migrating non-specific complex seen with tel-J did not form, however, consistent with the observation that unlabelled vsg-J could not compete this complex (Figure 1B). We found that nuclear extract derived from *L.tarentolae* promastigotes also contains a J-specific binding activity (Figure 2A). In this case, the bandshift resembles the *C.fasciculata* J-specific complex. Procyclic stage *T.brucei* do not contain J, so we tested nuclear extracts made from procyclic trypanosomes for J-binding activity. Using probe tel-J, procyclic extract gave non-specific complexes similar to bloodstream form extract, but the J-specific complex was absent (Figure 2A). Thus, of those cells tested, there is a firm correlation between the presence of J in DNA and the observation of J-binding activity in nuclear extract. To determine the nature of the J-binding activity, *T.brucei* bloodstream form and *C.fasciculata* nuclear extracts were pre-treated with proteinase K before addition to the binding reaction (data not shown). In both cases, this abolished complex formation, which was restored when phenylmethylsulfonyl fluoride (PMSF) was included in the proteinase K treatment step, indicating that the J-binding activity has a protein component (referred to as JBP).

J is made in two steps: first a thymine in DNA is converted to 5-(hydroxymethyl)uracil, which in turn is glucosylated to form J (Van Leeuwen *et al.*, 1998b). To determine whether the JBP found in trypanosome nuclear extract may have some affinity for the intermediate in J synthesis, we tested whether or not DNA containing

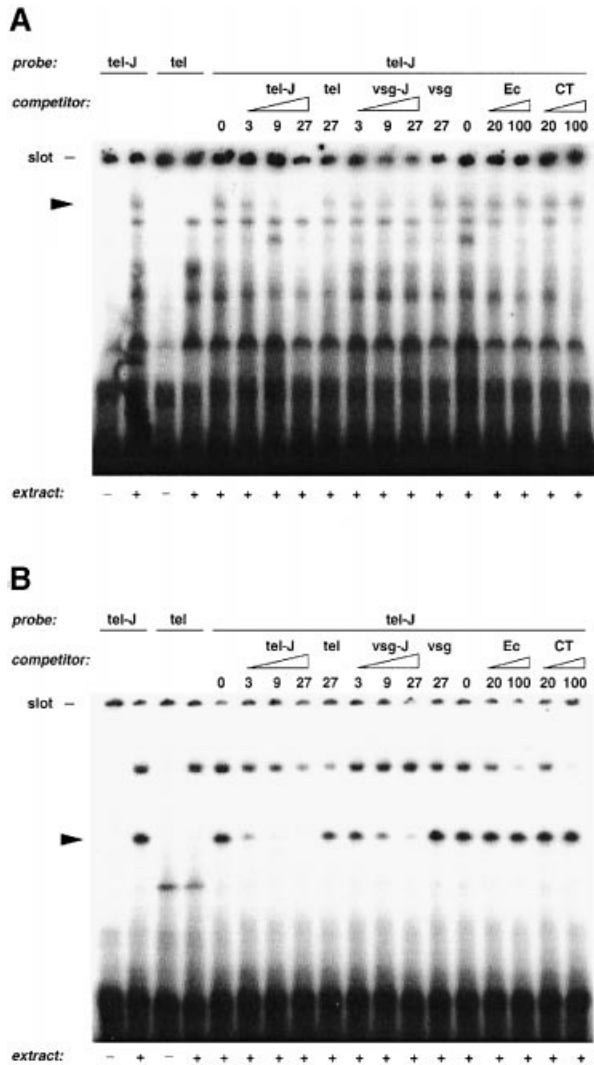


Fig. 1. Identification of a J-specific DNA-binding activity in trypanosomatid nuclear extracts. Bandshift assay of nuclear extract with telomere duplex DNA probes tel-J and tel (see Table I). Binding was competed by increasing molar excesses of various competitor DNAs, as indicated (Ec, *E. coli*; CT, calf thymus). +/–, presence or absence of nuclear extract in the reaction. The J-specific complex is indicated by an arrowhead. (A) *Trypanosoma brucei* bloodstream form nuclear extract. (B) *Crithidia fasciculata* nuclear extract.

5-(hydroxymethyl)uracil could inhibit JBP binding in the bandshift assay. First of all we found that bacteriophage ϕ e DNA, in which all thymines are replaced by 5-(hydroxymethyl)uracil, did not compete JBP binding even at 860-fold molar excess (data not shown; Table I). Furthermore, oligo duplex tel-V, which is the same as tel-J but has 5-(hydroxymethyl)uracil in place of J, had no effect on JBP complex formation (Figure 2B). We also tested pdJ and the nucleotide sugars UDPG and GDPG, none of which were found to compete the JBP complex (data not shown; Table I). These results suggest that JBP recognizes base J in DNA.

Purification of the *C. fasciculata* JBP

We concentrated on purifying the JBP from *C. fasciculata* as this seemed the most practical way to generate sufficient material for later analysis. Fractionation was monitored for JBP by bandshift assay. Precipitation of crude nuclear

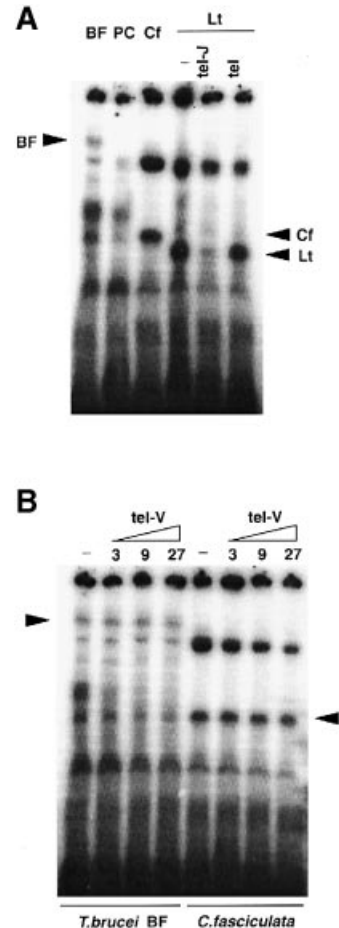


Fig. 2. J-specific binding of protein to DNA does not occur with nuclear extract from procyclic *T. brucei* and is not inhibited by hydroxymethyluracil-containing DNA. (A) Bandshift assay of kinetoplast nuclear extracts with duplex DNA probe tel-J. BF, *T. brucei* bloodstream form; PC, *T. brucei* procyclic; Cf, *C. fasciculata*; Lt, *L. tarentolae* promastigote. For *L. tarentolae*, binding reactions contained a 9-fold molar excess of competitor DNA tel-J or tel, as indicated. The J-specific complex for each species is indicated by an arrowhead. (B) Bandshift assay of *T. brucei* bloodstream form (BF) and *C. fasciculata* nuclear extracts with duplex DNA probe tel-J in the presence of increasing molar excesses of DNA competitor tel-V, the 5-(hydroxymethyl)uracil derivative of tel-J (see Table I). The *T. brucei* and *C. fasciculata* J-specific complexes are indicated by arrowheads.

extract with 20% polyethylene glycol (PEG) left most of the JBP in the supernatant and fractionated away the non-specific DNA-binding activity detected by probe tel-J (Figure 3A, lanes NE and P20 sup). Although the P20 sup fraction still contained a complex mixture of proteins as judged by SDS-PAGE and silver staining (Figure 3B), 80–85% of the bulk protein was removed routinely in this step. The P20 sup fraction was then passed over DNA-cellulose to remove any contaminating DNA-binding proteins and nucleases. JBP was recovered without loss in the flow-through (Figure 3A, lane DNA FT) and the total amount of protein in this fraction was unchanged. The final step in the purification scheme involved specific binding of JBP to J-containing DNA, using Dynabeads coated with streptavidin to immobilize biotinylated J-containing oligo duplex and retrieve JBP from the DNA FT fraction. We found that when duplex vsg-J containing a single J was used as ligand during affinity chromatography, JBP activity was absent from the unbound fraction

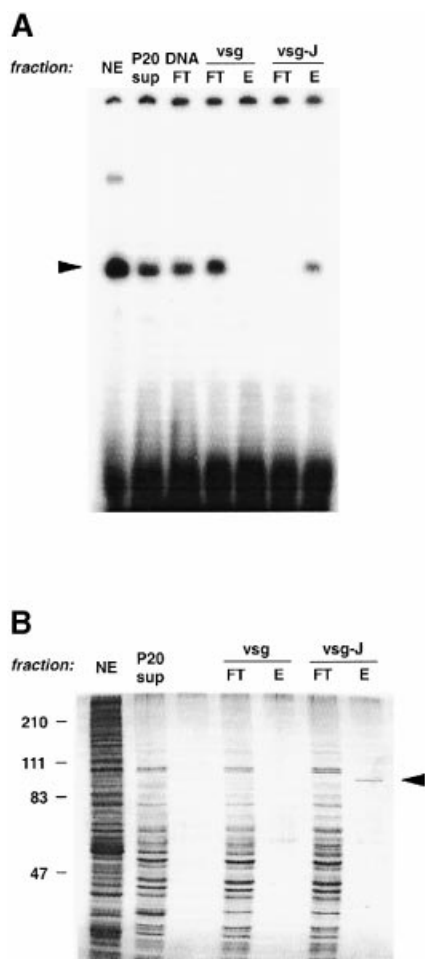


Fig. 3. Purification of the *C.fasciculata* JBP. JBP from *C.fasciculata* nuclear extract was purified through a combination of PEG precipitation and DNA affinity chromatography. (A) J-specific DNA-binding activity of the indicated fractions was assayed by bandshift with probe tel-J. One microgram of crude nuclear extract (NE) or the equivalent (subsequent fractions) was used per reaction. P20 sup, supernatant following PEG precipitation of crude nuclear extract; DNA FT, flow-through from DNA-cellulose chromatography of the P20 sup; vsg and vsg-J, DNA affinity chromatography of the DNA FT fraction using duplex oligos vsg or vsg-J as ligand; lanes FT and E are the DNA affinity chromatography flow-through (unbound) and eluate (bound), respectively. The J-specific complex is indicated by an arrowhead. (B) Protein analysis of the indicated fractions (as in A) by SDS-PAGE and silver staining. Two micrograms each of NE, P20 sup and DNA affinity chromatography FT were loaded on the gel. The amount loaded for each DNA affinity chromatography eluate came from an input of 160 μ g of DNA FT fraction. The migration of protein molecular mass standards (in kDa) is shown on the left. JBP is indicated by an arrowhead. Note that the weak bands of 60–70 kDa present in both eluate lanes arise from contaminants within the gel.

and reappeared following elution from the beads with high-salt buffer (Figure 3A, lanes vsg-J FT and E). JBP binding to the vsg-J-beads approached 100%, and we were able routinely to recover 50–90% JBP activity in the eluate. To control for non-specific binding to the oligo-beads, we also performed affinity chromatography using duplex vsg in place of vsg-J. This time all of the JBP activity remained in the unbound fraction, with nothing in the high-salt eluate (Figure 3A, lanes vsg FT and E). SDS-PAGE and silver staining of the eluate from the vsg-J affinity chromatography identified a single protein of ~90 kDa, which was absent from the control

(Figure 3B, lanes vsg-J E and vsg E, respectively). With protein from a larger scale purification, we were able to detect the 90 kDa JBP after staining with Coomassie Blue (data not shown). By comparison with protein molecular mass markers, we estimate that the yield of JBP is ~1 μ g per 2.5×10^{11} cells, an enrichment of ~50 000-fold.

Cloning the *C.fasciculata*, *T.brucei* and *L.tarentolae* JBP genes

Purified *C.fasciculata* JBP was analysed by SDS-PAGE, stained with silver, in-gel digested with trypsin and subjected to mass spectrometric sequencing. The peptide sequences obtained were then used to design degenerate DNA oligonucleotides to amplify the corresponding cDNA, which was used to clone the JBP gene by screening a size-fractionated *C.fasciculata* genomic library. Using this approach, we isolated a *C.fasciculata* gene with a predicted open reading frame of 811 amino acids, resulting in a protein with a molecular mass of 90 kDa (Figure 4). All seven of the peptide sequences obtained from the purified protein were found in the predicted JBP sequence.

To clone the *T.brucei* and *L.tarentolae* JBP genes by homology, we used a probe spanning amino acids 88–702 of the *C.fasciculata* JBP gene. Screening size-fractionated genomic libraries yielded a *T.brucei* gene and an *L.tarentolae* gene homologous to the *C.fasciculata* JBP gene (Figure 4). No other genes homologous to the *C.fasciculata* JBP gene were detected in these libraries. The *T.brucei* and *L.tarentolae* JBPs are predicted to contain 839 and 827 amino acids, respectively, with a molecular mass of 95 and 93 kDa, respectively. Alignment of the predicted protein sequences reveals that the *C.fasciculata* and *L.tarentolae* JBPs are more closely related to each other (64% identity) than they are to *T.brucei* (both 43% identity). However, all three proteins show significant blocks of similarity at the levels of amino acid sequence and predicted secondary structure, particularly in the N-terminal half. A computer-assisted search of the available databases revealed JBP to be unlike any other known protein.

Recombinant *C.fasciculata* and *T.brucei* JBPs bind specifically to J-containing DNA

To validate that the gene we had cloned really encodes the JBP, we produced full-length *C.fasciculata* JBP in *E.coli* using the pET expression system. Although almost 90% of the recombinant JBP (rJBP) was insoluble in the lysate pellet fraction, sufficient quantities of soluble rJBP could be recovered in the supernatant. The rJBP was purified by metal affinity chromatography by virtue of the N-terminal His₆ tag engineered on to the protein (Figure 5A, left panel). J-binding activity was then monitored by the bandshift assay. Figure 5A shows that rJBP formed a complex with probe tel-J, which migrated similarly to that observed with *C.fasciculata* nuclear extract. Furthermore, complex formation was competed only by J-containing DNA. Lysates from uninduced *E.coli* or from *E.coli* carrying the pET vector alone showed no J-binding activity (data not shown). These results prove that the gene we cloned is indeed the JBP gene.

We also expressed His-tagged recombinant *T.brucei* JBP in *E.coli*. In this case, expression was poor, resulting in amounts of soluble rJBP that could only be detected

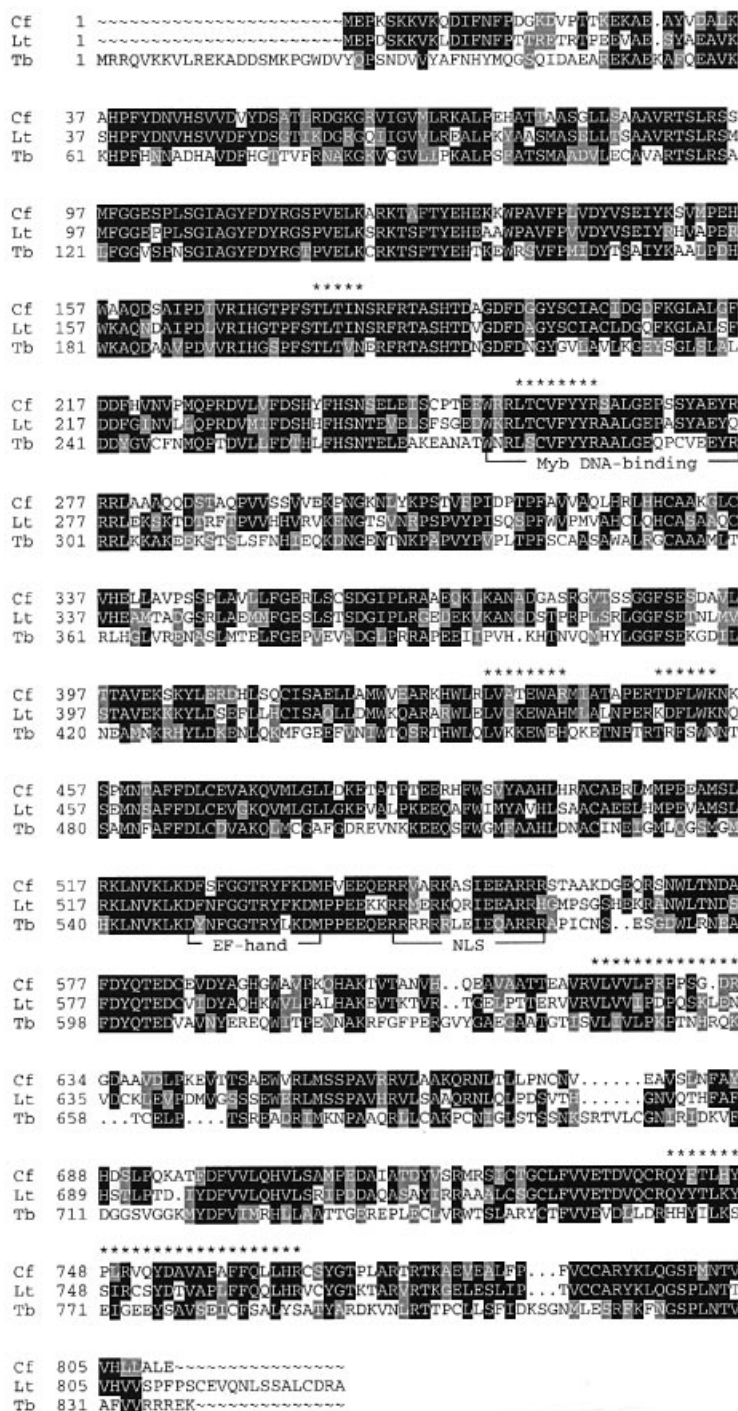


Fig. 4. Alignment of the amino acid sequences of the *C.fasciculata* (Cf), *L.tarentolae* (Lt) and *T.brucei* (Tb) JBPs. Amino acid identity is denoted by filled residues, similarity by shaded residues. Peptide sequences obtained from sequencing purified *C.fasciculata* JBP are indicated by asterisks above the sequence. Putative domains found by searching the PROSITE database are indicated below the sequence (Bairoch *et al.*, 1997): Myb DNA-binding, Myb DNA-binding domain repeat; EF-hand, EF-hand calcium-binding domain; NLS, bipartite nuclear localization signal.

by immunostaining with antibodies against the N-terminal His₆ tag (Figure 5B, left panel). Nevertheless, *T.brucei* rJBP was found to bind specifically to probe tel-J in the bandshift assay (Figure 5B, centre panel). Several complexes were formed, mostly migrating to a position in the gel similar to that seen for *C.fasciculata* JBP. This was unexpected since the J-specific complex we had characterized from *T.brucei* nuclear extract migrates much more slowly in the gel. The difference is unlikely to be

due to shorter forms of rJBP, as no significant degradation was observed following the DNA-binding reaction (Figure 5B, left panel). To confirm these results, we used a UV cross-linking assay to monitor J-specific binding and to detect the protein bound to J-containing DNA (Figure 5B, right panel). With *T.brucei* nuclear extract, a protein of ~110 kDa was cross-linked specifically to probe tel-J. After correction for the mass of the attached oligo duplex (end-labelled with ³²P), we estimate the size of

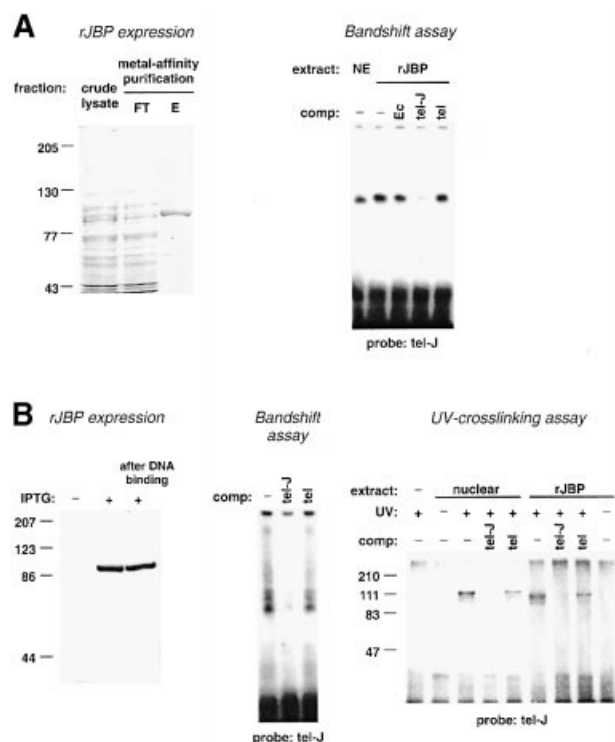


Fig. 5. Recombinant JBPs expressed in *E. coli* bind specifically to J-containing DNA. (A) Recombinant *C. fasciculata* JBP. The left panel shows the purification of *rJBP* expressed in *E. coli* and analysed by SDS-PAGE/Coomassie staining. Lanes show crude lysate of *E. coli* pET-CfJBP following induction (crude lysate), and the flow-through (FT) and eluate (E) fractions following metal affinity purification of the His₆-tagged *rJBP* from the crude lysate. The migration of protein molecular mass standards (in kDa) is shown on the left. The right panel shows J-specific DNA-binding activity of *rJBP* assayed by bandshift with probe tel-J and the indicated DNA competitors (comp) at 10-fold molar excess (Ec, *E. coli* DNA). NE, *T. brucei* bloodstream form nuclear extract. (B) Recombinant *T. brucei* JBP. The left panel shows expression of *rJBP* in *E. coli* analysed by Western blotting. His₆-tagged *rJBP* partially purified by metal affinity chromatography was detected by probing a protein blot with antibodies against the His₆ tag. Lanes show purified fractions from *E. coli* pET-TbJBP lysates before (–) and after (+) IPTG induction. Recombinant JBP recovered from a DNA-binding assay is also shown (lane ‘after DNA binding’). The migration of protein molecular mass standards (in kDa) is shown on the left. The centre panel shows J-specific DNA-binding activity of *rJBP* assayed by bandshift with probe tel-J and the indicated DNA competitors (comp) at 10-fold molar excess. The right panel shows J-specific DNA-binding activity of nuclear extract and *rJBP* assayed by UV cross-linking and analysed by SDS-PAGE/autoradiography. Lanes indicate the presence of DNA competitors at 10-fold molar excess (comp), UV irradiation (+/–) of the samples and the extract used in the binding reaction. The migration of protein molecular mass standards (in kDa) is shown on the left.

the J-binding-protein in nuclear extract to be ~97 kDa. This closely matches the predicted 95 kDa encoded by the *T. brucei* JBP gene we have cloned. We were also able to detect specific cross-linking of probe tel-J to the recombinant JBP (Figure 5B, left panel), confirming the J-binding activity of this protein.

Discussion

We have identified a J-binding protein in nuclear extracts of *T. brucei* bloodstream form and the related kinetoplastids *C. fasciculata* and *L. tarentolae*. Purification of the *C. fasciculata* JBP allowed the cloning of the

corresponding gene and its homologues from *T. brucei* and *L. tarentolae*. JBP binds only to double-stranded DNA containing the modified base J. DNA of the same nucleotide sequence but containing Thy instead of J is not recognized by JBP. In bloodstream form trypanosomes, J is found in such a variety of different sequence contexts that assigning a modification consensus site has not been possible (Van Leeuwen *et al.*, 1997). We tested two different sequences as substrate for JBP: the telomere repeat and a segment of a VSG gene. A large proportion of J is found in telomeric repeats and its location has been mapped previously (Van Leeuwen *et al.*, 1996), whilst the presence of J in the VSG gene has been inferred from a blocked restriction site (Van Leeuwen *et al.*, 1997). Both J-containing sequences bind the *T. brucei* and *C. fasciculata* JBPs, even though VSG genes are not found in *C. fasciculata*. By using two very different DNAs as probes for J binding, we focused on proteins that bind to J rather than to a particular nucleotide sequence. This was especially important with the telomere repeat, as proteins that bind to this sequence have already been identified in *T. brucei* (Eid and Sollner-Webb, 1995, 1997; Field and Field, 1996), *C. fasciculata* (Tzfati *et al.*, 1992) and *Leptomonas* (Field and Field, 1996). It is possible that some of the non-J-specific complexes we observed with the telomere repeat probe were due to these proteins or their homologues. None of these complexes appeared to be affected by the presence of J, so we did not investigate them any further. Although JBP can recognize base J in different sequences, it may do so with variable affinity. J-containing telomeric repeats compete JBP binding more effectively than J-containing VSG sequence, and the JBP–telomere repeat complex is resistant to high salt. This may be due to the differences in length, amount of J or sequence between the two duplex DNAs. Further experiments with a greater variety of J-containing sequences and purified recombinant protein, together with the determination of the crystal structure of JBP complexed to DNA, should resolve this issue. What is clear is that JBP does not bind the free nucleotide J, suggesting that J in the context of DNA is the target. This is reinforced by the fact that the glucose-free intermediate in J synthesis, 5-(hydroxymethyl)uracil in DNA, is also not bound by JBP.

If J has evolved as a tag on DNA to direct the binding of a specific protein, those organisms that contain J should also possess a JBP. We found this to be true for the three species we tested: *T. brucei*, *C. fasciculata* and *L. tarentolae*. In addition, no JBP activity was detected in procyclic form *T. brucei*, a stage in the life cycle where J is absent. Northern blot analysis indicates that the JBP mRNA is absent from procyclic *T. brucei* (data not shown). The J-specific complex we observed in the bandshift assay with *C. fasciculata* nuclear extract is most likely to be due to the binding of JBP alone to J-modified DNA, since recombinant *C. fasciculata* JBP expressed in *E. coli* produced the same complex. Although we cloned the JBP gene from *T. brucei* by virtue of its homology to the *C. fasciculata* gene, we believe that it encodes the J-binding activity we found in *T. brucei* nuclear extract for the following reasons. First, the encoded protein expressed in *E. coli* shows J-specific DNA binding in both bandshift and UV cross-linking assays. Secondly, the size of the J-binding protein in nuclear extract, as detected by

UV cross-linking, corresponds to that predicted from the gene. Thirdly, the J-specific, UV-cross-linked protein is absent from the nuclear extract of a JBP gene knockout cell line obtained very recently (M. Cross and P. Borst, unpublished results). Why the J-specific complex formed by endogenous JBP in nuclear extract differs from that formed by recombinant JBP expressed in *E. coli*, as judged by native gel electrophoresis, is unclear. It is possible that the *T. brucei* JBP has a tendency to aggregate and form multimers or bind other proteins, and that this might be somewhat less for the His-tagged protein than for the native JBP.

The amount of *C. fasciculata* JBP we were able to purify from nuclear extract was low. Given that we could recover >50% of the JBP present, as determined by bandshift assay, we estimate a yield of only 50 molecules of JBP per cell. As *C. fasciculata* contains $\sim 2.4 \times 10^5$ molecules of J per cell (Van Leeuwen *et al.*, 1998c), it is likely that we have succeeded in purifying only a fraction of the actual amount of JBP present. No JBP activity could be detected in the cytoplasmic fraction of the cell extract, suggesting that leakage of JBP during the isolation of nuclei is not a major problem (data not shown). Instead it is possible that nuclear extraction strips away only a subpopulation of JBP. The greatest proportion of JBP is expected at the telomeres as this is where most of J is located. Since we found that JBP binds to the telomere repeat probe even at a high salt concentration, it is possible that most of the JBP is not recovered during cellular extraction. Additional stabilization of JBP by interactions with telomere-binding proteins may also hinder extraction, similarly to what has been described in other organisms (Fang and Cech, 1995).

JBP is a novel protein and shows similarity throughout its sequence between *T. brucei*, *C. fasciculata* and *L. tarentolae*. Computer-assisted domain searching of the PROSITE database predicts several interesting regions conserved in all three JBPs (Figure 4). Based upon the *C. fasciculata* JBP sequence, amino acids 253–276 show 80% similarity to the Myb DNA-binding repeat signature (Biedenkapp *et al.*, 1988). Furthermore, all three JBPs contain conserved putatively helical regions proximal to this motif, as would be expected for a derivative of this helix–turn–helix-type DNA-binding domain. The central region of JBP is rich in aliphatic residues and shows 75% similarity to the leucine zipper motif (Landschulz *et al.*, 1999), and may therefore serve to recruit to J-modified DNA other proteins that could have eluded the nuclear extraction procedure we used. In the C-terminal half of JBP, amino acids 525–537 show 80% similarity to the EF-hand calcium-binding domain (Kawasaki and Kretsinger, 1995), and amino acids 544–556 contain the bipartite nuclear localization signal consensus sequence (Robbins *et al.*, 1991). We have no experimental data as yet to suggest that calcium is required for the function of JBP.

Previously, we have investigated the effect on *T. brucei* of altering the level of J in the genome through the use of thymidine analogues (Van Leeuwen *et al.*, 1998b). Lowering the J content caused a partial derepression of silent VSG gene expression site promoters. Raising the J content caused a reduction in the occurrence of chromosome breakage events that are sometimes associated with transcriptional switching between VSG gene expression

sites. We speculated that these effects are mediated by the packaging of J-containing DNA into a specialized chromatin structure via the recruitment for a protein factor. Our finding that base J directs the binding of a specific protein to DNA is consistent with this hypothesis. Little is known about chromatin structure in *T. brucei*, and it is a subject of debate with respect to VSG expression sites. Chromatin nuclease sensitivity assays to compare active and silent expression sites have given conflicting results (Greaves and Borst, 1987; Navarro and Cross, 1998). Recently, Navarro *et al.* (1999) showed that inactive expression site chromatin is accessible to bacteriophage T7 RNA polymerase. However, as the authors pointed out, this result does not rule out a role for chromatin modification in silencing expression sites. There are several well-documented cases in eukaryotes of epigenetic silencing mediated by DNA-binding proteins. Perhaps the closest analogy to base J is methylation of the 5-position of cytosine in CpG dinucleotides, a covalent DNA modification that is involved in the mitotic inheritance of transcriptionally repressed chromatin states (reviewed in Kass *et al.*, 1997). In vertebrates, differences in DNA methylation status are associated with gene imprinting, carcinogenesis and the suppression of mobile element activity (Jaenisch, 1997; Jones and Gonzalgo, 1997; Yoder *et al.*, 1997). A conserved family of proteins has been described that specifically bind methyl-CpG, and one of these, MeCP2, mediates transcriptional repression. Although there does not appear to be enough MeCP2 in the cell to occupy every methyl-CpG, it has been shown recently that MeCP2 recruits a co-repressor complex containing histone deacetylase activity (Jones *et al.*, 1998; Nan *et al.*, 1998). It now appears that methylated cytosine residues are important in guiding histone deacetylase to specific chromatin domains, and that the methylation-dependent transcriptional silencing relies upon histone deacetylation. It remains to be seen if the modified base J, like methylated cytosine, acts as the signal to target chromatin for histone deacetylation (or other modification) and subsequent inactivity in trypanosomes. The identification of a J-binding protein and the ability to produce large amounts of it in *E. coli* should allow us to address these questions.

Materials and methods

Cells and nuclear extracts

Trypanosoma brucei bloodstream form (clone 221a) and procyclic trypanosomes (strain 427) were grown as described (Van Leeuwen *et al.*, 1997). *Crithidia fasciculata* were grown in 3.7% brain-heart infusion broth containing 20 µg of haemin/ml. *Leishmania tarentolae* (tarVla) promastigotes were grown in SDM-79 (Brun and Schoenenberger, 1979). Nuclear extracts were prepared according to Field and Field (1996), except that the high-salt extracts were either dialysed against buffer H₁₀₀ [25 mM HEPES–NaOH pH 7.9, 100 mM KCl, 5 mM MgCl₂, 1 mM EDTA, 10% glycerol, 0.5 mM dithiothreitol (DTT)] or used directly in subsequent purification steps. The protein concentration of the nuclear extract was typically 2–4 mg/ml.

Bandshift assays

DNA oligonucleotides used in this study are listed in Table I. Oligos were end-labelled with [γ -³²P]ATP and purified by exclusion chromatography. For double-stranded DNA probes, the labelled oligo was annealed to its non-labelled complementary strand. Annealing was controlled by electrophoretic analysis of the duplex DNAs. Double-stranded oligo competitors were prepared in a similar manner except that both strands

were non-labelled. The *E.coli* and calf thymus DNA competitors were sonicated to fragments of ~500 bp and annealed.

Binding reactions contained 150 fmol of probe mixed with 1 µg of nuclear extract (or the equivalent for purified fractions) in 25 mM HEPES–NaOH pH 7.9, 1 mM EDTA, 1 mM DTT, 10 µg of bovine serum albumin (BSA), 2 µg of poly(dI–dC)–poly(dI–dC) (Sigma) and 4 µg of α-casein, in a final volume of 20 µl. KCl and MgCl₂ concentrations were adjusted to 80 and 2.5 mM, respectively, for *T.brucei* extracts, and 100 and 5 mM, respectively, for both *C.fasciculata* and *L.tarentolae* extracts. If used, competitor DNAs were added to the reaction before the probe. Reactions were incubated at 30°C for 30 min. Complexes were resolved by electrophoresis at 150 V through a native 4.5% polyacrylamide (19:1) gel in 0.5× TBE and detected by autoradiography.

Purification of *C.fasciculata* JBP

The *C.fasciculata* JBP purification scheme was based upon the fractionation of J-specific DNA-binding activity. Crude nuclear extract was precipitated with 20% PEG 6000 for 60 min on ice. Under these conditions, JBP remains in solution. After centrifugation at 15 000 g for 10 min, the supernatant was recovered, diluted with two volumes of buffer H₁₀₀ and dialysed against the same buffer, resulting in the PEG20 sup fraction. This was then passed over DNA cellulose (1:1 single-stranded to double-stranded calf thymus DNA; Pharmacia) equilibrated in buffer H₁₀₀ at a flow rate of 14 ml/h, and the flow-through collected (DNA FT fraction). JBP was then affinity purified from this fraction using a J-containing DNA oligo. Briefly, 25 pmol of a biotinylated version of duplex oligo vsg-J (see Table I) were coupled per milligram of Dynabeads–streptavidin (Dyna), according to the manufacturer. The oligo–beads were resuspended in buffer M (as buffer H₁₀₀ but containing no MgCl₂) and added to the DNA FT fraction at a concentration of 14 µg beads/µg of protein. EDTA was added to 5 mM, followed by an incubation on ice for 30 min. The beads were then washed three times for 20 min each with buffer M containing 0.32 mg/ml poly(dI–dC)–poly(dI–dC), the volume used being equal to the starting volume of the DNA FT fraction. JBP was then eluted from the beads by incubation on ice for 15 min with one volume of buffer M containing 500 mM KCl. The affinity-purified fractions were precipitated with 10% trichloroacetic acid (TCA) and analysed by SDS–PAGE and silver staining (Bio-Rad Silver Plus). Oligo–beads were regenerated for further use by incubation with 2 M KCl for 10 min and washing with buffer M.

Peptide sequencing

Purified protein was separated by SDS–PAGE and stained with silver. The protein was in-gel digested as described in Shevchenko *et al.* (1996) with the modification that the buffer contained 33% ¹⁸O-labelled water. This leads to an isotopic labelling to 33% of all cleaved peptides at the C-terminus. Peptides were extracted and analysed on an API III triple quadrupole tandem mass spectrometer (PE-Sciex, Toronto, Canada) equipped with a nanoelectrospray ion source as described (Shevchenko *et al.*, 1996; Neubauer and Mann, 1999). The peptides were sequenced *de novo* with mass spectrometry using a novel differential scanning technique. Briefly, two fragment spectra are acquired for each peptide, one from the ¹⁶O- and the ¹⁸O-labelled peptide in one spectrum, the second only from the ¹⁸O isotope. This can be done without reducing the ion transmission through the instrument by increasing the resolution of the selecting quadrupole filter to isotopic resolution (M.Wilm, G.Neubauer, L.Taylor, A.Shevchenko and A.Bachi, in preparation). The sequence of the peptides was determined comparing the two fragment spectra by identifying C-terminal fragment ions based on their different isotopic representation in both spectra.

Cloning of the *C.fasciculata*, *T.brucei* and *L.tarentolae* JBP genes

The following peptide sequence obtained from the purified *C.fasciculata* JBP was used successfully to design a degenerate oligonucleotide and clone a *C.fasciculata* cDNA: FTLHYPLR, corresponding to amino acids 743–750, was the basis for the complementary oligonucleotide 5'-GCGAATTCGGGTAGTGVAKSGTGAA-3' containing an *EcoRI* site for cloning. This primer together with a primer specific for the min-exon were used to amplify first-strand cDNA from *C.fasciculata* poly(A)⁺ RNA. A 1.8 kb *PstI* fragment from a cDNA with the correct predicted sequence was then used as a probe to make a genomic map of the *C.fasciculata* JBP locus and subsequently screen a size-fractionated genomic library for a 7.5 kb *HindIII*–*EcoRI* fragment containing the JBP gene.

The 1.8 kb *PstI* *C.fasciculata* JBP probe was then used to identify

homologous genes in *T.brucei* and *L.tarentolae*. For *T.brucei*, a size-fractionated genomic library containing 1.3 kb *HindIII* fragments was screened under conditions of low stringency (filters were washed at 3× SSC, 0.1% SDS at 42°C). The resulting positive clone contained only part of the *T.brucei* JBP gene, so the 1.3 kb *HindIII* fragment was used to probe a size-fractionated genomic library for a 7.5 kb *SacI* fragment containing the complete JBP gene. For *L.tarentolae*, the *C.fasciculata* JBP probe was used to screen a size-fractionated genomic library (filters were washed at 0.1× SSC, 0.1% SDS at 55°C) for an 8 kb *HindIII*–*ClaI* fragment containing the complete JBP gene.

For all JBP gene clones, convenient restriction fragments were subcloned and sequenced.

Expression of recombinant JBPs in *E.coli*

Recombinant full-length *C.fasciculata* and *T.brucei* JBPs were expressed in *E.coli* using the pET15b expression vector (Novagen). This system allows the production of recombinant protein containing an N-terminal His₆ tag to aid purification. *Escherichia coli* HMS 174(DE3)pLysS carrying the JBP expression plasmid were grown at 37°C to an A₆₀₀ of 0.6, then isopropyl-β-D-thiogalactopyranoside (IPTG) was added to 500 µM to induce protein expression and the cells allowed to grow for 7 h at 26°C. Cells were harvested by centrifugation, resuspended with lysis buffer (50 mM Tris–HCl pH 7.9, 500 mM NaCl, 5 µg/ml each of leupeptin and pepstatin) at 1/12.5 of the original culture volume, and frozen at –70°C. Following a 25°C thaw, the extract was passed several times through an 18 gauge needle. Cellular debris was removed by centrifugation at 10 000 g for 15 min at 4°C, and the resulting supernatant passed over 1 ml of TALON™ beads (Clontech) to purify the recombinant JBP, according to the manufacturer's instructions. Recombinant JBP was eluted with 50 mM imidazole. Approximately 5 ng of the resulting eluate were assayed for J-specific DNA-binding activity in the standard bandshift assay. Soluble *C.fasciculata* rJBP could be detected by Coomassie staining after SDS–PAGE, but soluble *T.brucei* rJBP could only be detected by Western blotting/ECL with antibodies against the His₆ tag.

UV cross-linking assay

J-specific DNA binding was also assayed by UV cross-linking. Binding reactions were carried out as described above, except that α-casein was omitted and 0.2 pmol of probe were used. A 3.8 µg aliquot of *T.brucei* bloodstream form nuclear extract or 10 ng of *T.brucei* recombinant JBP was used per reaction. After binding, reactions were irradiated on ice in a Stratagene UV Stratalinker at 0.5 J/cm². The samples were then boiled in loading buffer and proteins separated on an 8% polyacrylamide–SDS gel. Proteins cross-linked to the labelled DNA oligo were detected by autoradiography.

Accession numbers

The following sequences have been deposited in the DDBJ/EMBL/GenBank database: *T.brucei* JBP, accession No. AF182399; *C.fasciculata* JBP, accession No. AF182400; and *L.tarentolae* JBP, accession No. AF182401.

Acknowledgements

We would like to thank Inês Chaves, Dennis Dooijes, Herlinde Gerrits, Rainer Musmann and Ronald Plasterk for suggestions and critical reading of the manuscript, and Titia Sixma for help with protein structure predictions. M.C. was in part funded by an EC Marie Curie fellowship.

References

- Bairoch, A., Bucher, P. and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.*, **25**, 217–221.
- Bernards, A., van Harten-Loosbroek, N. and Borst, P. (1984) Modification of telomeric DNA in *Trypanosoma brucei*; a role in antigenic variation? *Nucleic Acids Res.*, **12**, 4153–4170.
- Biedenapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.H. (1988) Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature*, **335**, 835–837.
- Borst, P. and Van Leeuwen, F. (1997) β-D-Glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. *Mol. Biochem. Parasitol.*, **90**, 1–8.
- Brun, R. and Schoenenberger, M. (1979) Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Trop.*, **36**, 289–292.

- Colot, V. and Rossignol, J.L. (1999) Eukaryotic DNA methylation as an evolutionary device. *BioEssays*, **21**, 402–411.
- Eid, J.E. and Sollner-Webb, B. (1995) ST-1, a 39-kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29-base-pair subtelomeric repeat and its C-rich strand. *Mol. Cell. Biol.*, **15**, 389–397.
- Eid, J.E. and Sollner-Webb, B. (1997) ST-2, a telomere and subtelomere duplex and G-strand binding protein activity in *Trypanosoma brucei*. *J. Biol. Chem.*, **272**, 14927–14936.
- Fang, G. and Cech, T.R. (1995) Telomere proteins. In Blackburn, E.H. and Greider, C.W. (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 69–105.
- Field, H. and Field, M.C. (1996) *Leptomonas seymouri*, *Trypanosoma brucei*: a method for isolating trypanosomatid nuclear factors which bind *T.brucei* single-stranded g-rich telomere sequence. *Exp. Parasitol.*, **83**, 155–158.
- Gommers-Ampt, J.H., Van Leeuwen, F., de Beer, A.L., Vliegthart, J.F., Dizdaroğlu, M., Kowalak, J.A., Crain, P.F. and Borst, P. (1993) β -D-Glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T.brucei*. *Cell*, **75**, 1129–1136.
- Greaves, D.R. and Borst, P. (1987) *Trypanosoma brucei* variant-specific glycoprotein gene chromatin is sensitive to single-strand-specific endonuclease digestion. *J. Mol. Biol.*, **197**, 471–483.
- Grunstein, M. (1997) Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.*, **9**, 383–387.
- Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
- Jaenisch, R. (1997) DNA methylation and imprinting: why bother? *Trends Genet.*, **13**, 323–329.
- Jones, P.A. and Gonzalvo, M.L. (1997) Altered DNA methylation and genome instability: a new pathway to cancer? *Proc. Natl Acad. Sci. USA*, **94**, 2103–2105.
- Jones, P.L., Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.*, **19**, 187–191.
- Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) How does DNA methylation repress transcription? *Trends Genet.*, **13**, 444–449.
- Kawasaki, H. and Kretsinger, R.H. (1995) Calcium-binding proteins 1: EF-hands. *Protein Profile*, **2**, 297–490.
- Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1999) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759–1764.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905–914.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, **58**, 499–507.
- Nan, X., Ng, H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386–389.
- Navarro, M. and Cross, G.A.M. (1998) *In situ* analysis of variant surface glycoprotein expression-site promoter region in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **94**, 53–66.
- Navarro, M., Cross, G.A.M. and Wirtz, E. (1999) *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *EMBO J.*, **18**, 2265–2272.
- Neubauer, G. and Mann, M. (1999) Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal. Chem.*, **71**, 235–242.
- Pays, E., Delauw, M.F., Laurent, M. and Steinert, M. (1984) Possible DNA modification in GC dinucleotides of *Trypanosoma brucei* telomeric sequences; relationship with antigen gene transcription. *Nucleic Acids Res.*, **12**, 5235–5247.
- Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell*, **64**, 615–623.
- Sherman, J.M. and Pillus, L. (1997) An uncertain silence. *Trends Genet.*, **13**, 308–313.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins in silver-stained polyacrylamide gels. *Anal. Chem.*, **68**, 850–858.
- Tzfati, Y., Abieliovich, H., Kapeller, I. and Shlomai, J. (1992) A single-stranded DNA-binding protein from *Crithidia fasciculata* recognizes the nucleotide sequence at the origin of replication of kinetoplastid DNA minicircles. *Proc. Natl Acad. Sci. USA*, **89**, 6891–6895.
- Van Leeuwen, F., Wijsman, E.R., Kuyl-Yeheskiely, E., van der Marel, G.A., van Boom, J.H. and Borst, P. (1996) The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res.*, **24**, 2476–2482.
- Van Leeuwen, F., Wijsman, E.R., Kieft, R., van der Marel, G.A., Van Boom, J.H. and Borst, P. (1997) Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev.*, **11**, 3232–3241.
- Van Leeuwen, F., De Kort, M., van der Marel, G.A., Van Boom, J.H. and Borst, P. (1998a) The modified DNA base β -D-glucosyl-hydroxymethyluracil confers resistance to micrococcal nuclease and is incompletely recovered by 32 P-postlabeling. *Anal. Biochem.*, **258**, 223–229.
- Van Leeuwen, F., Kieft, R., Cross, M. and Borst, P. (1998b) Biosynthesis and function of the modified DNA base β -D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol. Cell. Biol.*, **18**, 5643–5651.
- Van Leeuwen, F., Taylor, M.C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R. and Borst, P. (1998c) β -D-Glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc. Natl Acad. Sci. USA*, **95**, 2366–2371.
- Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.

Received August 3, 1999; revised and accepted September 10, 1999